# Rendering the Sigmon Spectrum

Balancing Spectral Fidelity with Visual Aesthetics

Design & Signal-Processing Notes

June 16, 2026


**Abstract**

Sigmon is a real-time system-audio spectrum analyzer. Its display is not a raw power spectrum: a raw spectrum is numerically faithful but visually poor — it is noisy, bass-heavy, flickers violently, and clips on loud passages. This note documents the processing chain that turns the windowed FFT of the audio stream into the curve drawn on screen. The guiding principle throughout is that every transformation is a *monotone, locally information-preserving* reshaping of a quantity that is already on a logarithmic (decibel) axis: peaks are emphasized and noise is suppressed, but nothing that is present is deleted. We give the exact formulas and constants used in the shipping build.


## 1 Overview of the pipeline

Let $x[n]$ be the captured mono audio at sample rate $f_s$ (typically 48 000 Hz). One display frame is produced by the following stages, each described in its own section:

1. **Windowed FFT** → linear power per bin (§2).
2. **Log-frequency banding** by energy averaging (§3).
3. **Decibel conversion** and **perceptual tilt** (§4).
4. **Soft noise gate** near the floor (§5).
5. **Gaussian shaping** across bands (§6).
6. **Local-contrast enhancement** — the fidelity/aesthetic core (§7).
7. **Soft-knee limiting** toward the ceiling (§8).
8. **Temporal easing** and **peak-hold** (§9).
9. **Slew-limited adaptive floor** — the vertical scale (§10).
10. **Onset-driven "boom"** modulation (§11).

All shipping constants are collected in Table 1.

## 2 Windowed FFT and linear power

Each frame analyses the most recent $N = 8192$ samples. A larger window buys frequency resolution

$$\Delta f = \frac{f_s}{N} = \frac{48000}{8192} \approx 5.86\,\text{Hz},$$

which matters: at $N = 4096$ the lowest octaves are starved of bins and collapse into flat plateaus. The block is multiplied by a normalized Hann window

$$w[n] = \tfrac{1}{2}\left(1 - \cos\tfrac{2\pi n}{N-1}\right), \qquad n = 0, \ldots, N-1,$$

to suppress spectral leakage from the implicit rectangular truncation. A real-to-complex FFT then yields complex bins $X[k]$, from which we keep the *linear power* (magnitude squared), normalized by $N^2$:

$$P[k] \;=\; \frac{|X[k]|^2}{N^2}, \qquad k = 0, \ldots, \tfrac{N}{2} - 1, \qquad f_k = k\,\frac{f_s}{N}.$$

Working in power (not magnitude) is what makes the next step an honest *energy* average. Consecutive frames advance by a hop of $N/4$, i.e. 75% overlap, so the display updates smoothly without waiting a full window.

## 3 Log-frequency banding by energy averaging

The ear hears pitch logarithmically, so the $N/2$ linear bins are collapsed onto $B = 240$ bands spaced geometrically between $f_{\min} = 20\,\mathrm{Hz}$ and $f_{\max} = \min(16000,\, f_s/2)\,\mathrm{Hz}$. Band $b$ spans $[f_b, f_{b+1})$ with

$$f_b \;=\; f_{\min}\left(\frac{f_{\max}}{f_{\min}}\right)^{b/B}, \qquad f_c^{(b)} \;=\; \sqrt{f_b\, f_{b+1}} \quad \text{(geometric centre)}.$$

Let $\mathcal{I}_b = \{k : f_b \le f_k < f_{b+1}\}$ be the linear bins falling in the band. We average their *power*:

$$\bar{P}_b \;=\; \frac{1}{|\mathcal{I}_b|} \sum_{k \in \mathcal{I}_b} P[k].$$

Energy averaging (rather than averaging decibels, or taking a max) is the faithful choice: it conserves the total power landing in each band regardless of how many bins it contains, so wide high-frequency bands are not artificially inflated relative to narrow low-frequency ones.

## 4 Decibel conversion and perceptual tilt

Power is compressed to a decibel scale, with a small floor $\varepsilon = 10^{-12}$ guarding the logarithm:

$$D_b \;=\; 10\log_{10}\big(\bar{P}_b + \varepsilon\big).$$

Musical and natural signals carry most of their energy at low frequencies; a faithful flat-dB plot is therefore dominated by bass and looks lifeless up top. We counter this with a fixed *perceptual tilt* of $s = 3\,\mathrm{dB}$ per octave about a pivot $f_{\mathrm{ref}} = 1000\,\mathrm{Hz}$,

$$T_b \;=\; s\,\log_2 \frac{f_c^{(b)}}{f_{\mathrm{ref}}},$$

plus a constant lift $L = 6\,\mathrm{dB}$ so the curve sits comfortably in frame:

$$\boxed{V_b \;=\; D_b + T_b + L\,.}$$

The tilt is the same correction a “pink-noise reference” applies: pink noise, flat to the ear, becomes a flat line on the tilted axis. Crucially the tilt is a *deterministic function of frequency only* — it rescales the axis, it does not distort the relative heights of tones within an octave, so per-frequency fidelity is preserved.

## 5 Soft noise gate

To quiet the “sea” of low-level noise without hard-cutting it, values within a knee $g = 12\,\mathrm{dB}$ of the floor $F_0 = -70\,\mathrm{dB}$ are pulled down by a factor that vanishes at the floor. With $t = \mathrm{clip}_{[0,1]}\big((V_b - F_0)/g\big)$,

$$V_b' \;=\; \begin{cases} F_0 + (V_b - F_0)\,t, & V_b < F_0 + g, \\ V_b, & \text{otherwise.} \end{cases}$$

This is a smooth expander: content just above the floor is gently attenuated (quadratically, since both the offset and $t$ scale with $V_b - F_0$), while anything well above the floor is untouched. The baseline goes calm; real signal is preserved.

## 6 Gaussian shaping across bands

Banding leaves a jagged, pixel-noisy outline. A discrete Gaussian blur along the band axis (*not* the frequency–energy axis) rounds flat plateaus into smooth humps. With kernel half-width $r = \lceil 3\sigma \rceil$ and $\sigma_s = 1.4$,

$$g_\sigma[i] \;=\; \frac{1}{Z}\exp\Big(-\frac{i^2}{2\sigma^2}\Big), \quad i = -r, \ldots, r, \qquad Z = \sum_{i=-r}^{r} \exp\Big(-\frac{i^2}{2\sigma^2}\Big),$$

$$(S{*}g_\sigma)_b \;=\; \sum_{i=-r}^{r} g_\sigma[i]\, V'_{\mathrm{clip}(b+i)},$$

with edge clamping. Because $\sum_i g_\sigma[i] = 1$, the blur is mean-preserving: it relocates no energy between distant frequencies, it only softens the local outline. Call the result $\tilde{V}_b$.

## 7 Local-contrast enhancement: the fidelity/aesthetic core

This is the stage that makes a dominant motif "pop" above its neighbours while explicitly refusing to erase them. It is an *unsharp mask* on the band axis. Let $\tilde{V} * g_{\sigma_c}$ be a *wide* Gaussian blur of the already-shaped curve, with $\sigma_c = 6$ bands — this is the "local neighbourhood" reference. The detail (high-pass) component is

$$\delta_b \;=\; \tilde{V}_b - (\tilde{V} * g_{\sigma_c})_b,$$

which is positive exactly where a band rises above its surroundings. We add back a fraction $k = 0.75$ of it,

$$\boxed{C_b \;=\; \tilde{V}_b + k\,\delta_b, \quad \text{then clamp } C_b \;\geq\; \tilde{V}_b - A.}$$

The clamp with $A = 8\,\mathrm{dB}$ is the design promise made precise: a neighbour of a loud peak may be *reduced by at most* 8 dB, never removed. Two properties make this the right tool:

**Locality $\Rightarrow$ no bass flooding.**
The reference is a *local* average over $\sim\sigma_c$ bands, not the global maximum. A loud tone at 60 Hz and a quieter melody at 800 Hz are far apart on the log axis, so each is judged against its own neighbourhood. A loud bass region therefore cannot suppress a distant melodic peak — each peak is sharpened relative to its own surroundings.

**Sharpness-selectivity $\Rightarrow$ musicality.**
A narrow tonal spike has large $\delta_b$; a broad noise hump has small $\delta_b$. The enhancement thus favours *tonal* content (motifs) over broadband wash — "most distinctly tonal" rather than merely "loudest."

Equivalently, with $G_{\sigma_c}$ the blur operator, the unclamped map is the linear sharpening filter

$$C \;=\; \big(I + k(I - G_{\sigma_c})\big)\,\tilde{V} \;=\; \big((1+k)I - kG_{\sigma_c}\big)\,\tilde{V},$$

a textbook high-boost filter whose only nonlinearity is the per-band floor clamp.

## 8 Soft-knee limiting toward the ceiling

With a stabilized vertical scale (§10) a loud transient can still shoot past the 0 dB ceiling and clip into an ugly flat top. Rather than hard clip, the top $\rho = 6$ dB is compressed so the curve *asymptotically approaches* the ceiling. For $C_b$ above the knee $\theta = -6$ dB,

$$C_b^\star \;=\; \theta + \rho\Big(1 - e^{-(C_b-\theta)/\rho}\Big)\,, \qquad C_b > \theta,$$

and $C_b^\star = C_b$ below. As $C_b \to \infty$, $C_b^\star \to \theta + \rho = 0$ dB but never reaches it: peaks round off near the top instead of forming a plateau. The map is continuous and has unit slope at the knee ($\frac{d}{dC}C^\star\big|_{C=\theta} = 1$), so there is no visible kink where limiting begins. This $C_b^\star$ is the per-frame *target*, written $\hat{V}_b$ below.

## 9 Temporal easing and peak-hold

Frame-to-frame the displayed curve $Y_b$ chases the target $\hat{V}_b$ with an asymmetric one-pole filter (fast attack, slow release) at 60 fps:

$$Y_b \;\leftarrow\; Y_b + \big(\hat{V}_b - Y_b\big)\,\alpha_b, \qquad \alpha_b = \begin{cases} \alpha_\uparrow = 0.35, & \hat{V}_b > Y_b \quad \text{(attack)},\\ \alpha_\downarrow = 0.12, & \hat{V}_b \le Y_b \quad \text{(release)}.\end{cases}$$

A quiet band near the floor is settled faster ($\alpha_b \ge 0.45$) so the baseline stays still. The faint upper trace is a decaying *peak-hold*:

$$H_b \;\leftarrow\; \begin{cases} Y_b, & Y_b > H_b,\\ \max(F_0,\; H_b - \pi), & \text{otherwise},\end{cases} \qquad \pi = 0.35\,\text{dB/frame}.$$

Attack faster than release is the standard analyzer feel: onsets snap up, tails fall away gracefully.

## 10 Slew-limited adaptive floor

The vertical axis spans $[\,\Phi,\,0\,]$ dB, where the floor $\Phi$ adapts to overall loudness so quiet passages are not tiny and loud ones do not clip — but it must not *jump*, or the whole graph appears to lurch. Two safeguards:

A slow loudness envelope tracks the per-frame maximum $M = \max_b Y_b$ asymmetrically,

$$E \;\leftarrow\; E + (M - E)\,\beta, \qquad \beta = \begin{cases} 0.03, & M > E,\\ 0.012, & M \le E,\end{cases}$$

and the target floor sits a fixed dynamic range $R = 48$ dB below it (or is pinned to $\Phi = F_0$ when the user disables auto-ranging):

$$\Phi^{\text{tgt}} \;=\; \begin{cases} \max(F_0,\; E - R), & \text{auto-range on},\\ F_0, & \text{off}.\end{cases}$$

The floor is then moved toward that target under a hard *slew-rate limit* $\mu = 0.08$ dB per frame ($\approx 5$ dB/s):

$$\Phi \;\leftarrow\; \Phi + \operatorname{clip}_{[-\mu,\,\mu]}\big(\Phi^{\text{tgt}} - \Phi\big).$$

The slew limit is the key: regardless of how violently $\Phi^{\text{tgt}}$ moves, $\Phi$ can only *glide*. The soft limiter (§8) catches any transient that briefly outruns the gliding scale, so we obtain a steady axis *and* graceful peaks — the aesthetic middle ground between a fixed scale (which clips) and a fast adaptive one (which jumps). Finally each sample maps to screen height $h$ by

$$y_b \;=\; h\left(1 - \frac{\operatorname{clip}_{[\Phi,0]}(Y_b) - \Phi}{0 - \Phi}\right).$$

## 11 Onset-driven "boom" modulation

A light/shake accent fires on sharp rises in overall loudness. Let $\mu_t = \frac{1}{B}\sum_b Y_b$ be the frame's mean level and $\bar{\mu}$ a slow follower, $\bar{\mu} \leftarrow \bar{\mu} + (\mu_t - \bar{\mu})\,0.05$. The spectral-flux onset is $\phi = \mu_t - \bar{\mu}$. When it exceeds a threshold $\tau = 3\,\mathrm{dB}$ the boom is (re)triggered, then decays geometrically each frame:

$$\text{if } \phi > \tau: \quad \mathrm{b} \;=\; \min\Big(1,\; 0.35 + \tfrac{\phi-\tau}{\kappa}\Big), \qquad \mathrm{b} \;\leftarrow\; \mathrm{b}\cdot d,$$

with scale $\kappa = 6\,\mathrm{dB}$ and decay $d = 0.90$. The accent is purely presentational: a bloom of opacity $\propto \mathrm{b}$ in the palette accent colour, and a decaying vertical shake

$$\Delta y \;=\; \mathrm{b}\cdot A_s \cdot \sin(\varphi), \quad \varphi \mathrel{+}= 0.85 \text{ per frame}, \quad A_s = 5\,\mathrm{px},$$

applied to the trace only (the grid and labels never move). Because $\bar{\mu}$ chases $\mu_t$, sustained loudness stops re-triggering — the boom responds to *changes*, i.e. beats and hits, not steady level. A user intensity multiplier scales both bloom and shake; `Reduce Motion` forces $\Delta y = 0$ while keeping the bloom.

## 12 Why this is faithful

Every stage is one of three kinds of operation:

1. **Monotone axis maps** (dB conversion, frequency-only tilt, constant lift): order-preserving in amplitude, so relative tone heights are intact.
2. **Mean-preserving local reshaping** (energy banding, Gaussian shaping, unsharp contrast): no energy is moved between distant frequencies; emphasis is local and bounded (the $A = 8\,\mathrm{dB}$ clamp).
3. **Bounded display dynamics** (gate, soft limiter, easing, slew-limited floor): these shape *how* values are shown over time and near the rails, never inventing structure that the spectrum does not contain.

The aesthetic gains — a calm baseline, rounded musical peaks, a dominant motif that stands out, a steady axis, and a beat-synced accent — are thus obtained without sacrificing the property that a tall line means a loud frequency and a short line a quiet one. Fidelity and beauty are not traded; they are layered.

| **Symbol** | **Value** | **Meaning** |
|---|---|---|
| $N$ | 8192 | FFT window length (samples) |
| hop | $N/4$ | 75% overlap between frames |
| $B$ | 240 | display bands |
| $f_{\min}, f_{\max}$ | 20, 16000 Hz | band range |
| $\varepsilon$ | $10^{-12}$ | log-floor guard |
| $s,\ f_{\text{ref}}$ | 3 dB/oct, 1000 Hz | perceptual tilt |
| $L$ | 6 dB | constant lift |
| $F_0$ | $-70$ dB | absolute floor |
| $g$ | 12 dB | noise-gate knee |
| $\sigma_s$ | 1.4 bands | Gaussian shaping width |
| $k$ | 0.75 | local-contrast strength |
| $\sigma_c$ | 6 bands | contrast neighbourhood width |
| $A$ | 8 dB | max neighbour attenuation |
| $\theta,\ \rho$ | $-6$ dB, 6 dB | soft-limiter knee / headroom |
| $\alpha_{\uparrow}, \alpha_{\downarrow}$ | 0.35, 0.12 | attack / release |
| $\pi$ | 0.35 dB/frame | peak-hold decay |
| $R$ | 48 dB | adaptive dynamic range |
| $\beta$ | 0.03/0.012 | loudness-envelope rise/fall |
| $\mu$ | 0.08 dB/frame | floor slew limit ($\approx 5$ dB/s) |
| $\tau,\ \kappa,\ d$ | 3 dB, 6 dB, 0.90 | boom threshold / scale / decay |
| $A_s$ | 5 px | boom shake amplitude |

Table 1: Shipping constants for the rendering pipeline.